

**A SIMULATION MODEL OF MICROBIOLOGICAL
LABORATORY PROCEDURES**

Tommy Stephen Armstrong

LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIF. 94040

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THESIS

A SIMULATION MODEL OF
MICROBIOLOGICAL LABORATORY PROCEDURES

by

Tommy Stephen Armstrong

Thesis Advisor:

Hans J. Zweig

September 1973

Approved for public release; distribution unlimited.

T 1501 46

A Simulation Model
of
Microbiological Laboratory Procedures

by

Tommy Stephen Armstrong
Major, United States Army
B.S., Texas A&M University, 1962
D.V.M., Texas A&M University-
School of Veterinary Medicine, 1963
M.S., University of Maryland, 1968

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1973

ABSTRACT

Laboratory procedures, mathematical theory and distribution assumptions associated with two microbiological testing techniques are presented. A computer simulation model is then formulated and programmed based on these procedures, and thus the influences of changes in the number of microorganisms per sample, distribution of microorganisms within the sample, number of positive groups, probability of "false positives", distribution of "false positives" and technician analysis times are determined.

Using the basic simulation model as an experimental device, an example is presented to demonstrate its use in estimating the total time required to analyze a sample using each of the two procedures. Five variations of the basic model are presented to demonstrate the model's flexibility and sensitivity to fixing individual parameters.

Hypothesis testing is conducted on data obtained with the basic model and five variations. A significant Z value was obtained with variation two in which the probability of a false positive was set at zero. Results of all hypothesis testing are presented and a discussion of model data application in cost analysis is appended.

TABLE OF CONTENTS

I.	INTRODUCTION -----	4
II.	OBJECTIVES -----	6
III.	MPN ASSUMPTIONS AND THEORY -----	7
IV.	LABORATORY PROCEDURES -----	16
V.	THE SYSTEM TO BE MODELED -----	19
VI.	DESCRIPTION OF THE MODEL -----	21
VII.	OPERATION OF THE MODEL -----	23
VIII.	VARIATIONS OF THE MODEL -----	27
IX.	VERIFICATION OF RESULTS -----	28
X.	CONCLUSIONS -----	29
	TABLE I -----	31
	TABLE II -----	33
	APPENDIX 1 - Flow Chart -----	34
	APPENDIX 2 - Basic Program and Verification of Computational Procedures -----	35
	APPENDIX 3 - Basic Program -----	48
	APPENDIX 4 - First Variation -----	50
	APPENDIX 5 - Second Variation -----	52
	APPENDIX 6 - Third Variation -----	54
	APPENDIX 7 - Fourth Variation -----	56
	APPENDIX 8 - Fifth Variation -----	58
	APPENDIX 9 - Cost Analysis -----	60
	BIBLIOGRAPHY -----	83
	INITIAL DISTRIBUTION LIST -----	84
	FORM DD 1473 -----	85

I. INTRODUCTION

Laboratory microbiological analysis of animal origin food products for the determination of actual or potential health hazards is, at best, a cumbersome, time consuming and expensive procedure for which no perfect alternative is likely to be found in the near future.

Further, because it is impractical, if not impossible, to examine samples for all potentially pathogenic microorganisms, laboratory methods currently in use rely heavily upon the isolation and identification of members of "indicator" groups.

Briefly, the rationale for using "indicator" groups is that they are readily and reliably cultured in the laboratory and are fairly good predictors of general microbiological quality. (1)

Among the most widely used "indicator" groups is that which comprises the coliform organisms. These organisms are primarily members of the family Enterobacteriaceae, and the two genera Escherichia and Aerobacter supply the majority of the strains. The American Public Health Association defines the group as "---all aerobic and facultative anaerobic, gram-negative, non-sporeforming rods capable of fermenting lactose with the production of acid and gas at 32 degrees to 35 degrees centigrade within 48 hours incubation on solid or in liquid media." Included in this broad

grouping are some strains of the genera Klebsiella, Paracolobactrum, Erwinia and Serratia, as well as the Escherichia and Aerobacter.

Food specifications require that products meet standards based in some instances on total coliform counts. Other specifications stipulate limits for the genus Escherichia while still others have become more stringent and now require that food products contain no members of those E. Coli varieties most commonly associated with the intestinal tracts of man and other vertebrates.

Laboratories responsible for analyzing products under these specifications are required to perform one or more of the standard coliform procedures designed to enumerate the total coliform population of the product under examination. (One of these standard procedures will be discussed at length in the next section of this paper.) In addition, laboratories must perform specific identification procedures on E. Coli varieties to determine whether they are of the type for which a zero tolerance has been established.

While the total coliform procedures are fairly well standardized and must be adhered to rigorously by all laboratories, there are optional techniques available for performing the E. Coli typing. Laboratories operating under personnel, time and budgetary constraints would therefore derive substantial benefit from selecting those analytical techniques which were most efficient in terms of resource utilization and, at the same time, provide an acceptable degree of reliability.

In general, because of the large number of variables involved in these laboratory techniques, a straightforward analytic solution to the question of which procedure is most efficient in a particular laboratory is not available to the laboratory supervisor. Further, because of the time, expense and laboratory facilities required to perform these procedures, many laboratories can't conduct the additional testing necessary to arrive at a satisfactory solution to the question on an experimental basis.

II. OBJECTIVES

The primary objective of this paper is to develop and demonstrate the use of an analytic procedure for evaluating the relative efficiency of two microbiological laboratory methods. Specifically, the microbiological methods to be considered are coliform serotyping techniques associated with "Most Probable Number (MPN)" coliform determinations.

The basic analytic tool to be employed in this analysis is a computer simulation model. A simulation model was chosen because, as Naylor (2) states, simulation techniques allow us to conduct situational experiments that would ordinarily be too expensive and/or too cumbersome to perform physically. Clearly, the laboratory procedures to be modeled fit both categories.

Secondary objectives associated with the procedures to be modeled and the computer simulation to be demonstrated are:

1. To present MPN theory and to describe related laboratory procedures in sufficient detail for development of the model.

2. To discuss the specific system to be modeled.

3. To describe the model and variations of the model.

4. To conduct hypothesis testing on total analysis time data obtained with the model and to discuss conclusions drawn from these results.

Finally, Appendix 9 of this paper will consider the general subject of cost analysis as it relates to laboratory procedures of this type and, in particular, will discuss the application of data obtained with the basic model to the question of dollar cost efficiency.

III. MPN ASSUMPTIONS AND THEORY

The standard "Most Probable Number" (MPN) Coliform procedure forms the basis for the techniques to be modeled and analyzed. Therefore, a clear understanding of the assumptions and theory of MPN determinations is essential to the interpretation and application of the model to be presented.

A. ASSUMPTIONS

There are two principal assumptions. In statistical language, the first is that the organisms are distributed randomly (uniformly) throughout the sample. This means that an organism is equally likely to be found in any part of the sample, and that there is no tendency for pairs or

groups of organisms either to cluster together or to repel one another. In practice this implies that the sample is thoroughly mixed, and if the volume is not too great some mechanical device is employed for this purpose. This will be discussed further in the "laboratory procedures" section of this paper.

The second assumption is that each subsample from the sample, when incubated in the proper culture medium, is certain to exhibit growth whenever the subsample contains one or more organisms. This will be discussed further in the "model assumptions" section under "false positives". Also, if the culture medium is poor, or if there are factors which inhibit growth, or if the presence of more than one organism is necessary to initiate growth, the MPN gives an underestimate of the true sample density.

B. THEORY

Mathematically, MPN theory relates the probability that there will be no growth in a subsample to the density of organisms in the original sample. Suppose that the sample contains V ml., the subsample contains v ml., and that there are actually b organisms in the sample. By the second assumption, there will be no growth if and only if the sample contains no organisms. (Disregard the possibility of false positives for the moment.) Then, calculate the probability that none of these b organisms is in the subsample.

Consider a single organism. By the first assumption, the probability that it lies in the sample is simply the ratio of the volume of the subsample to that of the original sample, i.e. v/V . The probability that it is not in the subsample is therefore $(1 - v/V)$. Since there is assumed to be no kind of attraction or repulsion between organisms, these two probabilities hold for any organism, irrespective of the positions of the other organisms. (Strictly, this requires the additional assumption that the space occupied by an organism is negligible relative to v .) Consequently, by the multiplication theorem in probability, the probability that none of the b organisms is in the sample is

$$p = (1 - v/V)^b$$

When v/V is small, this is closely approximated by

$$p = e^{-vb/V}$$

where e is the base of natural logarithms. Finally, since b/V is the density S of organisms per ml., we have

$$p = e^{-vS}$$

where p is the probability that the subsample is sterile.

Consider the case of a single dilution. If n subsamples, each of volume v , are taken, and if s of these are found to be sterile, the proportion s/n of sterile samples is an estimate of p . Hence we obtain an estimate d of the density S by the equation

$$\frac{s}{n} = e^{-vd}$$

This gives

$$d = -\frac{1}{v} \ln \left(\frac{s}{n} \right) = -\frac{2.303}{v} \log \left(\frac{s}{n} \right)$$

where \ln and \log stand for logarithms to base e and to base ten respectively.

The estimate d is the "most probable number" of organisms per ml. of the original sample.

In this case, the concept of MPN is scarcely needed. It becomes useful, however, in the more complex situations where several dilutions are used.

If p is the probability that a sample is sterile, the probability that s out of n samples are sterile is given by the binomial distribution as

$$\frac{n!}{s! (n-s)!} p^s (1-p)^{n-s}$$

Since $p = e^{-vS}$, this expression may be written as

$$\frac{n!}{s! (n-s)!} e^{-svS} (1-e^{-vS})^{n-s}$$

If we have obtained s sterile samples out of n , this formula enables us to plot the probability of this event against the true density S . Such curves always have a single maximum.

A curve of this type suggests a method for estimating S , for if we are considering two possible values of S , it seems reasonable to prefer the one which gives a higher probability to the result that was actually observed. This argument, carried to its conclusion, leads to a choice of S for which the probability of obtaining the observed result is greatest. It is this value of S that is called the "most probable number" of organisms in the original sample.

In practice, more than one dilution is usually needed. The reason is that the precision of the mpn is very poor when the volume v in the subsample is such that the subsamples are likely to be all fertile or all sterile. When all are fertile, the maximum on the probability curve occurs when S is infinite, so that the estimated density is infinite. When all are sterile the estimated density is zero, as may be verified from the equations above. Thus a single dilution is successful only if v happens to be chosen so that some samples are sterile and some are fertile. Such a choice of v can be made only if the density S is known fairly closely in advance. As a practical matter, S is not known in advance. In default of this knowledge, the practice is to use several dilutions in the hope that at least one of them will give some sterile and some fertile subsamples.

To illustrate the general problem, consider the case of three dilutions. Let the suffix i indicate the dilution. For the i^{th} dilution the volume of subsample is v_i , and s_i out of n_i samples are found to be sterile. How do we estimate S from these results?

From above we can obtain a separate estimate for each dilution

$$d_i = - \frac{2.303}{v_i} \log \left(\frac{s_i}{n_i} \right)$$

However, the best way to combine the three estimates d_i into a single value is not obvious. Since, as we have seen, some dilutions give very poor estimates, it is not satisfactory to take the arithmetic mean.

One solution is provided by the MPN concept which extends easily to this situation. Following the approach used in the previous section, first write down the probability of obtaining the observed results for any hypothetical value of the true density S . The observed results are that s_1 samples out of n_1 are sterile at the first dilution, s_2 out of n_2 at the second and s_3 out of n_3 at the third. The probability that these three events should all happen is the product of three terms. As before, the graph of this probability against S shows a single maximum. The value of S at this maximum is taken as the MPN.

The value of the MPN cannot be written down explicitly. The equation it satisfies is as follows: (3)

$$s_1 v_1 + s_2 v_2 + s_3 v_3 = \frac{(n_1 - s_1) v_1 e^{-v_1 d}}{1 - e^{-v_1 d}} + \frac{(n_2 - s_2) v_2 e^{-v_2 d}}{1 - e^{-v_2 d}} + \frac{(n_3 - s_3) v_3 e^{-v_3 d}}{1 - e^{-v_3 d}}$$

In laboratories where the numbers of subsamples n_i and the dilution ratios are standardized, it is convenient to have a table which gives the MPN for all sets of results that are likely to occur. (4)

In the procedure to be modeled, we will only consider the case of three dilutions and five subsamples per dilution.

Although the number of dilutions and replications within dilutions is standardized by laboratory operating procedures for most specification testing, an understanding of the rationale for selecting dilution and replication numbers is useful in those instances when a sample is expected to contain an unusual level of contamination.

Generally, in preparation for an estimation by the MPN procedure, three decisions must be made as follows:

1. What range of sample volume is to be examined.
2. What dilution factor is to be used.
3. How many subsamples (replications) should be taken per dilution.

These decisions must in some way be related to a prior knowledge of the limits within which the true level of microbiological contamination is likely to lie and on the precision required in the estimate obtained by this procedure. Specifically, it follows from the previous discussion that the best estimate will be obtained from volumes of sample in which it is unlikely that all replicates will be fertile or that all replicates will be sterile. Then, in a series of dilutions, the expected number of contaminants in the highest sample volume selected for testing should be at least one. Otherwise, there is a risk that all samples will be sterile. Similarly, the expected number of contaminants in the lowest sample volume should not exceed two in order to avoid an unreasonable risk that all replicates will be fertile. Using this line of thought, the dilution series will be able to estimate any density of contamination that lies between $1/\text{Highest Volume}$ and $2/\text{Lowest Volume}$.

This rule is satisfactory if a sizeable number of replications (twenty or more) are being taken at each dilution. With small sample replicate numbers (five or less) which are required in the procedure we are discussing due to

time and expense of large replicate numbers, the above generalization is too lenient in that it allows too great a risk that all replicates will be fertile. Suppose, as in our example, that we have three ten fold dilutions with sample volumes 1/100, 1/10 and 1/1. By the generalization above, we should be able to estimate densities between 1 and 200 microorganisms per ml. If, on the other hand, the true density of microorganisms in the sample happens to be 200 per ml., so that the expected number of microorganisms per replication in the lowest sample volume is two, then the probability of a sterile sample at this dilution is e^{-2} or, 0.135. The probability of a fertile sample is then (1 - probability of a sterile sample) or (1 - 0.135 = 0.865). Then, if five replicates are used per dilution as in our case, the probability that all are fertile is 0.865^5 , or 0.484. Clearly, at the two higher concentrations all samples are very likely to be fertile. Thus we have at best a fifty-fifty chance that all samples (replicates) will be fertile which necessitates rerunning the sample at other dilutions to obtain a satisfactory estimate. On the other hand, if laboratory procedures permit and the expense is not too great, it might be well to consider larger numbers of replicates. For example, if twenty replicates were used, the probability that all are fertile becomes $(0.865)^{20}$, or only about 0.05.

The lesson to be learned from this is that it is safer to reduce the upper density when the number of replicates

per dilution must be small. In practice, the upper density is reduced from 2/vol to 1/vol. This is used by first guessing or estimating from existing laboratory records, the two limits between which we can be reasonably certain that the true microbiological density lies. The sample volumes are then chosen so that the volume of the highest density is greater than or equal to 1/lowest estimate of true density. Similarly, the volume of the lowest density is chosen to be less than or equal to 1/highest estimate of the density. For example, if we are confident that the density is somewhere between a low of 10 and a high of 750 per ml., the highest sample volume should be at least 1/10 ml.. Similarly, the lowest sample volume should not be more than 1/750 ml.. In this example, as in our case, three ten fold dilutions 1/10, 1/100, 1/1000 would amply cover this range of densities. This range of densities is standardized for most applications in microbiological laboratory testing and there is no real advantage to considering a different dilution ratio. As stated by Cochran (5), "if the total number of samples (replications) in the whole series is kept fixed, the average precision is practically the same for any dilution ratio between two and ten."

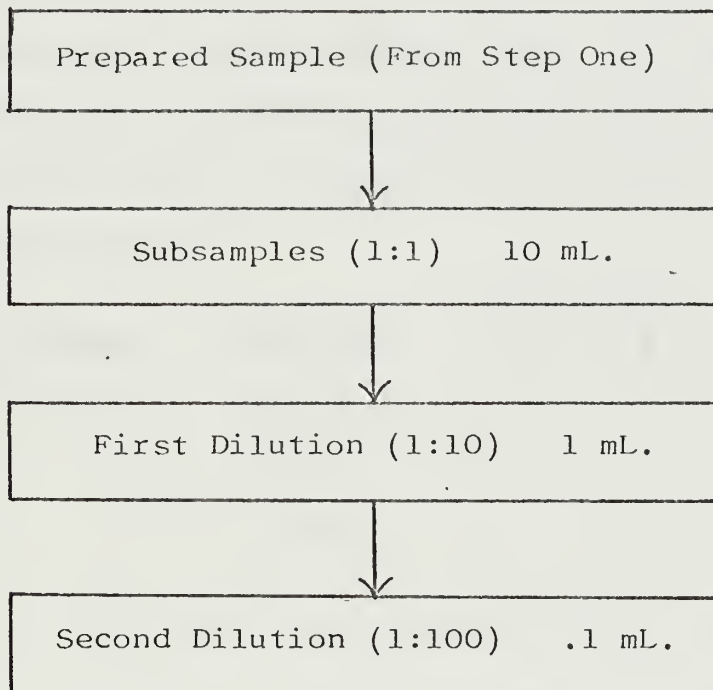
Thus, in routine testing, the recommended procedure of using three ten fold dilutions and five replicates per dilution has proven to be the most useful combination and for that reason, results are tabulated (see Table 1). An example of the use of this table will be presented in the next section.

IV. LABORATORY PROCEDURES

Consider a sample submitted for MPN Coliform and E. coli typing. This sample would be processed as follows:

1. The sample would be thoroughly mixed with a measured volume of diluent in an attempt to achieve the uniformity of organism distribution assumed by the MPN procedure.

2. Five subsamples are selected and diluted as shown in the following schematic:



3. Subsamples and dilutions are inoculated into appropriate growth media.

4. Inoculated subsamples and dilutions are incubated for twenty-four hours.

5. At the end of 24 hours, subsamples any of whose dilutions are positive are transferred to confirmatory media and/or are examined individually for E. coli type.

6. Those confirmatory subsamples which were transferred are examined at the end of an additional 24 hours incubation at $45.5 \pm .2$ degrees C. If positive at this point, they are confirmatory for E. coli.

7. Individual subsamples may now be examined for E. coli type. Negative subsamples are observed again at the end of 48 hours and if negative then they are discarded.

Results from this laboratory procedure are normally recorded in matrix form as follows: (Rows are dilutions and columns are replicates.)

<u>Sample Number</u>	<u>Dilution</u>	<u>Tube Number</u>				
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>
1	1:1	+	+	-	-	+
	1:10	+	+	-	-	-
	1:100	+	-	-	-	-

Each plus in the matrix represents a tube in which growth is observed and each minus represents a tube in which no growth is observed. If these results are from confirmatory tubes, the MPN per 100 milliliters may be obtained from the MPN table (see Table 1).

Tabular values are related to the MPN values per gram of the sample as follows:

Consider a sample in which one gram of solid matter is suspended in ten milliliters of liquid. In step one above, suppose that the sample is diluted ten fold (that is, sample is mixed with diluent on a one in ten basis). Then, following step one our testing dilution contains one gram per hundred milliliters liquid volume. In this example, the MPN per gram can be read directly from the table. Our sample matrix shows three positive tubes in the 1:1 dilution, two positive tubes in the 1:10 dilution and one positive tube in the 1:100 dilution. Then, reading from the table under the 3-2-1 values gives an MPN per 100 ml. of 17.

Clearly, if the original dilution represents something other than one gram in 100 ml. of liquid, tabular results must be adjusted. This is easily accomplished by the following formula:

$$\frac{\text{MPN from table}}{100} \times \begin{array}{l} \text{dilution factor} \\ \text{of middle} \\ \text{tube in series} \end{array} = \text{MPN per gram}$$

V. THE SYSTEM TO BE MODELED

The system to be modeled is that part of the analysis which requires that the positive subsamples (replicates) be examined individually for E. coli type. As discussed in the laboratory procedures section, this typing may be accomplished in two basic ways.

A. PROCEDURE A

At step seven in the laboratory procedure the technician selects those sample fermentation tubes which show gas (carbon dioxide) production. Each positive tube is then further examined for E. coli type by a macroagglutination procedure in which the E. coli contaminant acts as the antigenic agent and illicitly an agglutination of the type specific antisera in one of the ten typing tubes to be implanted. If the contaminant is not E. coli, no specific agglutination will be illicitly from the antisera in the ten typing tubes and it may be concluded that the contaminant was not E. coli or, more generally, that the fermentation tube had shown gas production due to any one or more of a wide variety of nonspecific causes all of which will be treated under the general classification "false positive". It will be noted that a false positive required exactly as much technician time to examine as did the tubes in which E. coli was present. In terms of resource utilization, this procedure can result in fewer total serotype tubes implanted

and examined and if the number of positive confirmatory tubes is small there may be a significant saving of technician time.

B. PROCEDURE B

At step five in the laboratory procedure the technician can implant ten subgroup (serotype) tubes at the same time the confirmatory E. coli tubes are being implanted. This routine offers the advantage of saving technician time during the implanting procedure but clearly requires that the technician implant a large number of tubes for each sample (50 tubes per sample). Samples for analysis will be generated by the model on the basis of distribution assumptions in the MPN procedure. Individual technician times, numbers of contaminants per sample, and the occurrence of false positives are arbitrarily established for demonstration purposes only. All parameters in this system except those related to the basic MPN assumptions could be easily and quickly determined in the laboratory prior to application of the model for a specific laboratory procedure.

In order to make this model as general as possible, positive tubes within a dilution are referred to as antigenic groups. Similarly, positive serotypes within a group are referred to as antigenic subgroups. Further, rather than restrict the nomenclature in the model to coliform groups, all organisms in a sample are referred to as microbiological contaminants. Hopefully, these generalities will encourage readers to examine the possibility of applying the model to a variety of laboratory procedures.

VI. DESCRIPTION OF THE MODEL

A. FLOW CHART

A flow chart of the program is attached as appendix 1.

B. EXPLANATION OF PROGRAM LISTING

- A&MATRIX - Represent the sample to be analyzed. The five rows of the matrix represent the five replicates (subsamples) which are referred to as Antigenic Groups and the ten columns represent the serotype tubes referred to as Antigenic Subgroups.
- K - Counter used in the program to keep track of the number of samples analyzed.
- M - Counter to determine the number of microbiological contaminants entered in the sample matrix.
- N - Number of samples to be analyzed.
- IX,KX,MX - Seed values for the random number generator.
- LA - Calculated time required for a technician to analyze one sample using procedure A.
- LB - Calculated time required for a technician to analyze one sample using procedure B.
- NAT - Random time required for analysis of one replicate (group) using procedure A.
- NBT - Random time required for analysis of one group using procedure B.
- LAS,LBS - Square of LA and LB.
- UMLAS - Sum of squares of LA.
- UMLBS - Sum of squares of LB.

NT - Number of microbiological contaminants in a sample.

NG - Number of positive replicates (groups) in the confirmatory MPN tubes.

RX - A uniformly distributed random variable from 0 to 1.

IROW - A random group to be included in the sample.

JCOL - A random subgroup to be included in the sample.

TIMEA - Sum of analysis times for procedure A.

TIMEB - Sum of analysis times for procedure B.

TTIMEA - Mean of analysis times for procedure A.

TTIMEB - Mean of analysis times for procedure B.

BTIMEA - Variance of analysis times for procedure A.

BTIMEB - Variance of analysis times for procedure B.

CTIMEA - 95% lower confidence limit of mean for procedure A.

CTIMEB - 95% lower confidence limit of mean for procedure B.

DTIMEA - 95% upper confidence limit of mean for procedure A.

DTIMEB - 95% upper confidence limit of mean for procedure B.

QTIMEA - Standard deviation of analysis times for procedure A / \sqrt{N} .

QTIMEB - Standard deviation of analysis times for procedure B / \sqrt{N} .

ZSTAT - Calculated Z value for testing the null hypothesis of no difference between mean analysis times for the two procedures.

VII. OPERATION OF THE SIMULATION MODEL

A matrix of sample contaminants is generated and printed out as follows:

		Antigenic Subgroup									
		<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>
Antigenic Group	<u>1</u>	0	0	0	0	1	0	0	0	1	0
	<u>2</u>	1	0	1	0	0	0	0	0	0	0
	<u>3</u>	0	0	0	0	0	0	0	0	0	0
	<u>4</u>	0	0	1	1	0	0	0	0	0	0
	<u>5</u>	1	0	0	0	0	0	1	0	0	0

Where the 1's indicate that a contaminant is present and the 0's indicate that no contaminant is present. As stated earlier, the antigenic groups 1 thru 5 correspond to the five subsamples (replications) prepared for the MPN procedure and the antigenic subgroups correspond to the ten possible (hypothetical) serotypes of the microbiological contaminant. Random variables for these entries are generated by the simulation model based on the assumption of normality in organism distribution from the MPN theory.

The computer first generates a random variable for matrix row (group) and then generates a random variable for matrix column (subgroup). These two numbers identify the specific tube in which a microbiological contaminant will be entered. The computer then scans the matrix (sample) and enters a 1 in the proper row and column. If a 1 has

previously been entered in that matrix row and column, the computer generates a new random variable for matrix row and a new random variable for matrix column and repeats the above process until the matrix (sample) contains the specified number of microbiological contaminants.

The computer then counts and records the numbers of positive groups (including false positives) in each generated sample, prints it out, computes technician times for the sample by each of the two procedures and calculates statistics on means, variances, confidence intervals and Z values for means according to the following scheme:

\bar{X} = Sample Mean

μ = Population Mean

s^2 = Sample Variance

s = Sample Standard Deviation

σ^2 = Population Variance

σ = Population Standard Deviation

Theory - For large N (by the central limit theorem)

$$\frac{\sqrt{N} (\bar{X} - \mu)}{\sigma} \approx N(0,1)$$

$$\text{then, } P(-1.96 \leq \frac{\sqrt{N} (\bar{X} - \mu)}{\sigma} \leq 1.96) = .95$$

and, using s^2 as an estimate for σ^2 this becomes

$$P(\bar{X} - 1.96 \frac{s}{\sqrt{N}} \leq \mu \leq \bar{X} + 1.96 \frac{s}{\sqrt{N}}) = .95$$

for the 95% confidence interval about the sample mean (\bar{X}).

The model computes the values by keeping a running sum of total times for each procedure (TIMEA and TIMEB), a running sum of squares of total times (UMLAS and UMLBS) and number of samples processed (N). After completing all sample processing, the model computes sample means (\bar{X}) by dividing TIMEA and TIMEB by N.

Sample variances are computed by the equation

$$S^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N-1}$$

For computational convenience and because of large N in the exercise, this is computed in the model by

$$S^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{N}}{N-1}$$

$$S^2 \approx \frac{\sum X_i^2}{N} - \frac{(\sum X_i)^2}{N} \cdot \frac{1}{N}$$

$$S^2 \approx \frac{\sum X_i^2}{N} - \left(\frac{\sum X_i}{N}\right)^2$$

then, from the values calculated by the model for the above:

$$BTIMEA = \frac{UMLAS}{N} - (TTIMEA)^2$$

and, similarly

$$BTIMEB = \frac{UMLBS}{N} - (TTIMEB)^2$$

then, factors for 95% confidence limits are computed

$$QTIMEA = \frac{\sqrt{BTIMEA}}{\sqrt{N}}$$

similarly,

$$QTIMEB = \frac{\sqrt{BTIMEB}}{\sqrt{N}}$$

The hypothesis testing for differences between means is conducted as follows:

\bar{X}_1 and \bar{X}_2 are the sample means obtained from large sample of size N drawn from populations having means μ_1 and μ_2 and standard deviations σ_1 and σ_2 . Then we can test the hypothesis of no difference between means ($\mu_1 = \mu_2$) using the statistic

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma(\bar{X}_1 - \bar{X}_2)}$$

where

$$\sigma(\bar{X}_1 - \bar{X}_2) \approx \sqrt{\frac{s_1^2 + s_2^2}{N}}$$

Here, the Z statistic is used rather than the t statistic because of the large sample size (400). In the model, the Z statistic is computed as

$$ZSTAT = \frac{TTIMEA - TTIMEB}{\sqrt{\frac{BTIMEA + BTIMEB}{N}}}$$

Then, referring to the Normal probability tables, for a two tail test and .05 level of significance:

1. If the calculated Z value is greater than 1.96 or less than -1.96, reject the hypothesis.

2. If the calculated Z value is less than 1.96 and greater than -1.96, accept the hypothesis.

See Table 2 for a summary of results obtained with the basic model and five variations in which one or more of the variables is fixed (held constant). These variations will be described in the next section and will be discussed individually in Appendices 4 - 8.

VIII. VARIATIONS OF THE MODEL

Five variations of the basic model were used in order to demonstrate the flexibility of the model and the overall change in results due to fixing individual variables. In each variation, the random number process is unaltered by the process of fixing a variable.

The five variations are as follows:

1. The number of contaminants (NT in the computer program listing) was fixed. (Appendix 4)

2. The probability of a false positive was set at zero. (Appendix 5)

3. The analysis time for technician on procedure A was fixed at seven minutes per positive group. (Appendix 6)

4. The analysis time for technician on procedure B was fixed at seven minutes per positive group. (Appendix 7)

5. Both technician times were fixed. (Appendix 8)

IX. VERIFICATION OF RESULTS

Verification of results obtained with the basic model and the five variations was accomplished manually as follows:

1. In order to verify the individual sample matrices, an initial run using a sample size of twenty, in which the basic model prints out each sample matrix number, the complete matrix, the identity and number of groups, false positives, analysis times for each sample and procedure is attached as Appendix 2. The entries in each matrix were verified by counting them individually and comparing the results with those tabulated by the computer following each sample. (See table in Appendix 2)

2. Confidence limits were verified manually by computing the results individually as shown in the following example.

For the basic model - Procedure A - Appendix 3

$$N = 400 \quad \bar{X} = 34.97 \quad S^2 = 80.567$$

$$\text{Then, } 95\% \text{ C.I.} = 34.97 \pm 1.96 \frac{\sqrt{80.567}}{\sqrt{400}}$$

$$\text{Upper C.I.} = 34.97 + .878$$

$$\approx 35.848$$

$$\text{Lower C.I.} = 34.97 - .878$$

$$\approx 34.092$$

Rounding these gives the values in Table 2 and in Appendix 3.

3. Z values were verified manually as shown in the following example.

For the basic model - Appendix 3

$$\begin{aligned}
 Z &= \frac{\bar{X}_1 + \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{N}}} \\
 &= \frac{34.973 - 34.937}{\sqrt{\frac{80.568 + 50.808}{400}}} \\
 &\approx \frac{.035}{\sqrt{.328}} \\
 &\approx .061
 \end{aligned}$$

Computer value from Table 2 (and from Appendix 3) = .06105.

X. CONCLUSIONS

Results obtained with the basic model and the five variations are summarized in Table 2. Conclusions based on these results are as follows:

1. For the basic model and all five variations, it must be concluded that the true population mean analysis times lie between the 95% confidence limits shown in the table unless a one in twenty sampling error has been made.

2. For the basic model and variations 1, 3, 4 and 5, the hypothesis of no difference between mean analysis times must be accepted. Or, stated another way, we must conclude that the observed differences between mean analysis times

for the two simulated procedures is due to chance alone at this level of significance.

3. For variation 2, the hypothesis of no difference between mean analysis times must be rejected. Thus we may conclude with 95% confidence that there is a real difference between mean analysis times, and, because the Z value is negative, that procedure A is significantly better than procedure B. In fact, referring to the Normal probability tables, it can be seen that with a Z value this large, our confidence in this conclusion can exceed 99%. Having obtained a Z value this large with variation 2, the laboratory supervisor might well pursue the question of false positives further by performing a sensitivity analysis on the range of probabilities from 0 to .2 and thereby identify the specific level of false positives necessary to produce a statistically significant difference between the two simulated procedures. That is, find the probability level for false positives at which the Z value no longer exceeds 1.96. (See Appendix 5)

In summary, it must be recalled that all parameter assignment in the preceeding example was arbitrary and that conclusions based on these hypothetical values are not intended to imply that Procedure A is, in general, better than Procedure B.

TABLE I

Most Probable Numbers Per 100 ml. of Sample, Planting
5 Portions in each of 3 Dilutions in Geometric Series

Positives with				Positives with				Positives with			
10 ml.	1 ml.	0.1 ml.	MPN	10 ml.	1 ml.	0.1 ml.	MPN	10 ml.	1 ml.	0.1 ml.	MPN
0	0	0	...	1	0	0	2.0	2	0	0	4.5
0	0	1	1.8	1	0	1	4.0	2	0	1	6.8
0	0	2	3.6	1	0	2	6.0	2	0	2	9.1
0	0	3	5.4	1	0	3	8.0	2	0	3	12
0	0	4	7.2	1	0	4	10	2	0	4	14
0	0	5	9.0	1	0	5	12	2	0	5	16
0	1	0	1.8	1	1	0	4.0	2	1	0	6.8
0	1	1	3.6	1	1	1	6.1	2	1	1	9.2
0	1	2	5.5	1	1	2	8.1	2	1	2	12
0	1	3	7.3	1	1	3	10	2	1	3	14
0	1	4	9.1	1	1	4	12	2	1	4	17
0	1	5	11	1	1	5	14	2	1	5	19
0	2	0	3.7	1	2	0	6.1	2	2	0	9.3
0	2	1	5.5	1	2	1	8.2	2	2	1	12
0	2	2	7.4	1	2	2	10	2	2	2	14
0	2	3	9.2	1	2	3	12	2	2	3	17
0	2	4	11	1	2	4	15	2	2	4	19
0	2	5	13	1	2	5	17	2	2	5	22
0	3	0	5.6	1	3	0	8.3	2	3	0	12
0	3	1	7.4	1	3	1	10	2	3	1	14
0	3	2	9.3	1	3	2	13	2	3	2	17
0	3	3	11	1	3	3	15	2	3	3	20
0	3	4	13	1	3	4	17	2	3	4	22
0	3	5	15	1	3	5	19	2	3	5	25
0	4	0	7.5	1	4	0	11	2	4	0	15
0	4	1	9.4	1	4	1	13	2	4	1	17
0	4	2	11	1	4	2	15	2	4	2	20
0	4	3	13	1	4	3	17	2	4	3	23
0	4	4	15	1	4	4	19	2	4	4	25
0	4	5	17	1	4	5	22	2	4	5	28
0	5	0	9.4	1	5	0	13	2	5	0	17
0	5	1	11	1	5	1	15	2	5	1	20
0	5	2	13	1	5	2	17	2	5	2	23
0	5	3	15	1	5	3	19	2	5	3	26
0	5	4	17	1	5	4	22	2	5	4	29
0	5	5	19	1	5	5	24	2	5	5	32

TABLE I (Continued)

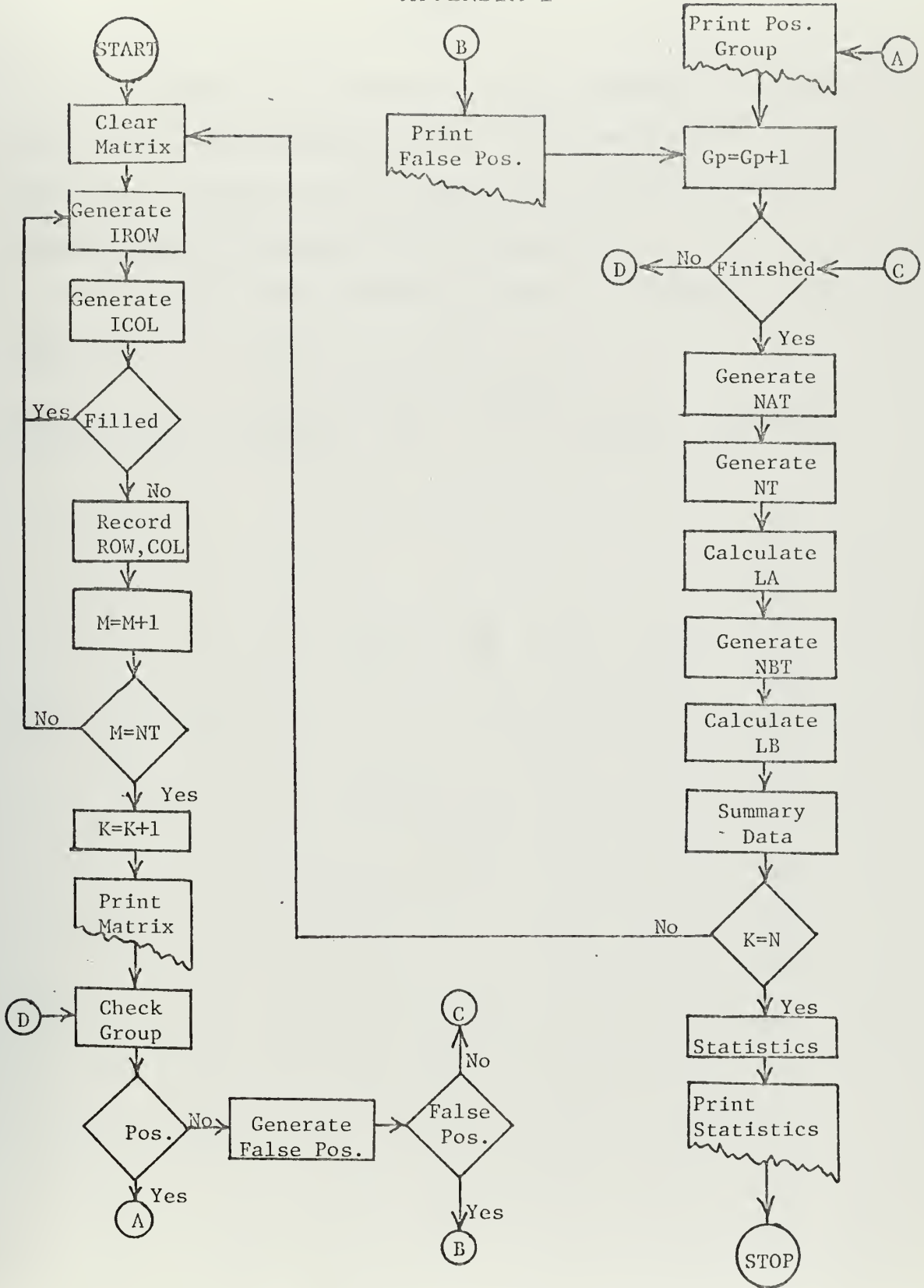
Most Probable Numbers Per 100 ml. of Sample, Planting
5 Portions in each of 3 Dilutions in Geometric Series

Positives with				Positives with				Positives with			
10 ml.	1 ml.	0.1 ml.	MPN	10 ml.	1 ml.	0.1 ml.	MPN	10 ml.	1 ml.	0.1 ml.	MPN
3	0	0	7.8	4	0	0	13	5	0	0	23
3	0	1	11	4	0	1	17	5	0	1	31
3	0	2	13	4	0	2	21	5	0	2	43
3	0	3	16	4	0	3	25	5	0	3	58
3	0	4	20	4	0	4	30	5	0	4	76
3	0	5	23	4	0	5	36	5	0	5	95
3	1	0	11	4	1	0	17	5	1	0	33
3	1	1	14	4	1	1	21	5	1	1	46
3	1	2	17	4	1	2	26	5	1	2	64
3	1	3	20	4	1	3	31	5	1	3	84
3	1	4	23	4	1	4	36	5	1	4	110
3	1	5	27	4	1	5	42	5	1	5	130
3	2	0	14	4	2	0	22	5	2	0	49
3	2	1	17	4	2	1	26	5	2	1	70
3	2	2	20	4	2	2	32	5	2	2	95
3	2	3	24	4	2	3	38	5	2	3	120
3	2	4	27	4	2	4	44	5	2	4	150
3	2	5	31	4	2	5	50	5	2	5	180
3	3	0	17	4	3	0	27	5	3	0	79
3	3	1	21	4	3	1	33	5	3	1	110
3	3	2	24	4	3	2	39	5	3	2	140
3	3	3	28	4	3	3	45	5	3	3	180
3	3	4	31	4	3	4	52	5	3	4	210
3	3	5	35	4	3	5	59	5	3	5	250
3	4	0	21	4	4	0	34	5	4	0	130
3	4	1	24	4	4	1	40	5	4	1	170
3	4	2	28	4	4	2	47	5	4	2	220
3	4	3	32	4	4	3	54	5	4	3	280
3	4	4	36	4	4	4	62	5	4	4	350
3	4	5	40	4	4	5	69	5	4	5	430
3	5	0	25	4	5	0	41	5	5	0	240
3	5	1	29	4	5	1	48	5	5	1	350
3	5	2	32	4	5	2	56	5	5	2	540
3	5	3	37	4	5	3	64	5	5	3	920
3	5	4	41	4	5	4	72	5	5	4	1600
3	5	5	45	4	5	5	81				

TABLE 2
Summary of Means and Z Values

<u>Model</u>	<u>Procedure</u>	<u>Mean</u>	95% Confidence Limits		<u>Z Value</u>	<u>Conclusion</u>
			<u>Lower</u>	<u>Upper</u>		
Basic	A	34.97	34.09	35.85	0.06	Accept
	B	34.94	34.24	35.64		
Var. 1	A	35.49	34.82	36.16	0.28	Accept
	B	35.35	34.65	36.05		
Var. 2	A	31.85	31.05	32.65	-5.71	Reject
	B	34.94	34.24	35.64		
Var. 3	A	34.79	34.04	35.55	-0.27	Accept
	B	34.94	34.24	35.64		
Var. 4	A	34.97	34.09	35.85	-0.06	Accept
	B	35.00	35.00	35.00		
Var. 5	A	34.79	34.04	35.55	-0.54	Accept
	B	35.00	35.00	35.00		

APPENDIX 1



APPENDIX 2

This appendix is included for the purpose of displaying the basic fortran program used in this model and to illustrate the procedure used to manually verify the model.

Verification of Computational Procedures

Individual samples shown on pages through of this appendix are counted and listed below:

<u>Sample Number</u>	<u>Positive Groups</u>		<u>Deviation</u>
	<u>Computer Count</u>	<u>Manual Count</u>	
1	3	3	0
2	3	3	0
3	2	2	0
4	3	3	0
5	2	2	0
6	1	1	0
7	1	1	0
8	3	3	0
9	2	2	0
10	1	1	0
11	3	3	0
12	2	2	0
13	4	4	0
14	1	1	0
15	3	3	0
16	1	1	0
17	1	1	0
18	3	3	0
19	2	2	0
20	3	3	0

Thus, it is readily seen that there is no difference between manual counts of positive groups and computer counts. Further, Z statistics can be verified manually from results shown in Appendices 3-8.

Consider the data in Appendix 3 for example:

$$\begin{aligned}
 Z &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{N}}} \\
 &= \frac{34.97249 - 34.93750}{\sqrt{\frac{80.56763 + 50.80859}{400}}} \\
 &= \frac{.03499}{\sqrt{.328}} \\
 &= \frac{.03499}{.5727} \\
 &\approx .061
 \end{aligned}$$

Computer Value = .06105


```

IMPLICIT INTEGER*2(A,M),INTEGER(P,I)
DIMENSION A(5,10),B(2,5),MATRIX(5,10)
EQUIVALENCE (A,MATRIX)
K=0
LA=0
LB=0
NT=20
NT=3
TIMEA=0
TIMEB=0
UNLAS=0
UNLBS=0
MX=12493
KX=8*MX+3
IX=1331
CATA B/ 'GP-O', 'NE ', 'GP-T', 'WG ', 'GP-T', 'FREE', 'GP-F', 'CUR ',
'GP-F', 'IVE '
C CLEAR THE MATRIX AND SET ALL VALUES TO ZERO
C
100 DC 1051 I=1,5
DC 1052 J=1,10
A(I,J)=0
1052 CCATINUE
1051 M=0
C GENERATE A RANDOM VARIABLE FOR GROUP (MATRIX RCh)
C
1053 IX=IX*KX
RX=0.5+FLOAT(IX)*2.328306E-10
IRCW=RX*5+1
C GENERATE A RANDOM VARIABLE FOR SUBGROUP (MATRIX COLUMN)
C
IX=IX*KX
RX=0.5+FLOAT(IX)*2.328306E-10
JCOL=RX*10+1
C CHECK TO SEE IF GROUP (I), SUBGROUP (J) IS FILLED
C
IF (MATRIX(IROW,JCOL).EQ.1)GO TO 1053
M=M+1
MATRIX(IROW,JCOL)=1
C CHECK TO SEE IF SAMPLE CONTAINS NT TYPES OF MICROORGANISMS
C
IF (M.EQ.NT)GO TO 1054

```



```

GC TO 1053
C INCREMENT SAMPLE NUMBER
C 1054 K=K+1
C WRITE(6,9500)K
C 9500 FORMAT('0','SAMPLE NUMBER',1X,15)
C PRINT OUT THE SAMPLE MATRIX
C 9501 WRITE(6,9501)((A(I,J),J=1,10),I=1,5)
C 9501 FORMAT('0',10I2//',',10I2//',',10I2//',',10I2)
C EXAMINE MATRIX A (SAMPLE) FOR POSITIVES
C NG=0
C DO 1000 I=1,5
C 1000 J=1,10
C IF (A(I,J).EQ.0)GO TO 500
C COUNT THE NUMBER OF POSITIVE GROUPS
C NG=NG+1
C PRINT GROUPS IN WHICH AGGLUTINATION IS OBSERVED
C 9530 WRITE(6,9530)(B(P,I),P=1,2)
C 9530 FORMAT('0',2A4//',',AGGLUTINATION IN ',2A4)
C GC TO 1000
C 500 CCNTINUE
C GENERATE A RANDOM VARIABLE FOR 'FALSE POSITIVES'
C IX=IX*KX
C RX=0.5+FLOAT(IX)*2.328306E-10
C IF (RX.GT.0.2)GO TO 1000
C ADD ONE TO GROUPS POSITIVE
C NG=NG+1
C PRINT GROUP IN WHICH FALSE POSITIVE OCCURS
C 9540 WRITE(6,9540)(B(P,I),P=1,2)
C 9540 FORMAT('0',2A4//',',FALSE POSITIVE IN ',2A4)
C 1000 CCNTINUE
C GENERATE A RANDOM VARIABLE FOR TECHNICIAN

```



```

C      IX=IX*KX
      RX=C.5+FLOAT(IX)*2.328306E-10
      NAT=5*RX+5
      NT=5*RX+1
C
C      CALCULATE PROCESSING TIME
C
C      LA=15+NG*NAT
C
C      GENERATE A RANDOM VARIABLE FOR TECHNICIAN
C
C      IX=IX*KX
      RX=C.5+FLOAT(IX)*2.328306E-10
      NBT=5*RX+5
C
C      CALCULATE PROCESSING TIME
C
C      LB=5*NBT
      WRITE(6,9550)LA,LB
9550  FORMAT(10,'TIME TC PERFORM PROCEDURE A WAS ',I3,I3,'MINUTES',/
10,'TIME TC PERFORM PROCEDURE B WAS ',I3,I3,'MINUTES')
C
C      KEEP A RUNNING SUM OF TOTAL TIME FOR EACH PROCEDURE
C
C      TIMEA=TIMEA+LA
      LAS=LA*LA
      UMLAS=UMLAS+LAS
      TIMEB=TIMEB+LB
      LES=LB*LB
      UMLBS=UMLBS+LBS
C
C      CHECK THE NUMBER OF SAMPLES COMPLETED
C      IF LESS THAN N MAKE THE NEXT SAMPLE
C
C      IF (K.LT.N)GO TO 100
C
C      AFTER COMPLETING ALL N SAMPLES, COMPUTE STATISTICS FOR EACH PROCEDURE
C
      TTIMEA=TIMEA/FLOAT(N)
      TTIMEB=TIMEB/FLOAT(N)
      BTIMEA=UMLAS/FLOAT(N)-TTIMEA*TTIMEA
      BTIMEB=UMLBS/FLOAT(N)-TTIMEB*TTIMEB
      ZSTAT=(TTIMEA-TTIMEB)/SQRT((BTIMEA*BTIMEB)/FLCAT(N))
      GTIMEA=SQRT(BTIMEB)/SQRT(FLOAT(N))
      GTIMEB=SQRT(BTIMEA-1.96*QTIMEA
      CTIMEA=TTIMEA+1.96*QTIMEA

```



```

CTIMEB=TTIMEB-1.96*QTIMEB
CTIMEB=TTIMEB+1.96*QTIMEB

C PRINT STATISTICS ON PROCESSING TIMES
C
9510 WRITE(6,9510)N,TTIMEA,CTIMEA,DTIMEA,BTIMEA
    FORMAT(11,'RESULTS OF ANALYSIS TIME WAS',F8.5,/,
    2CNFIDENCE LIMIT',F8.5,/,95%,F8.5,/,
    3',VARIANCE OF ANALYSIS TIME',F8.5,/)
9520 WRITE(6,9520)N,TTIMEB,CTIMEB,DTIMEB,BTIMEB
    FORMAT(11,'RESULTS OF ANALYSIS TIME WAS',F8.5,/,
    2CNFIDENCE LIMIT',F8.5,/,95%,F8.5,/,
    3',VARIANCE OF ANALYSIS TIME',F8.5,/)
9521 WRITE(6,9521)ZSTAT,STATISTIC FOR MEANS IS ',F8.5)
    FORMAT(11,'ZSTAT',F8.5,/,
    2CNFIDENCE LIMIT',F8.5,/,
    3',VARIANCE OF ANALYSIS TIME',F8.5,/)
END

```


SAMPLE NUMBER 1

O C O O O 1 O C O O
O O O O O O O O O O
C 1 O O O O C C O O
C C O 1 O O O O O O
C C O O C O C C O O

AGGLUTINATION IN GP-ONE

AGGLUTINATION IN GP-THREE

AGGLUTINATION IN GP-FOUR

TIME TO PERFORM PROCEDURE A WAS 36 MINUTES

TIME TO PERFORM PROCEDURE B WAS 40 MINUTES

SAMPLE NUMBER 2

C C C C O C C C O 1
O C O O O O O O O O
O C O C O O C C O 1
O C O O O O O O O O
O C O O O O C 1 O O

AGGLUTINATION IN GP-ONE

FALSE POSITIVE IN GP-TWO

AGGLUTINATION IN GP-THREE

AGGLUTINATION IN GP-FIVE

TIME TO PERFORM PROCEDURE A WAS 43 MINUTES

TIME TO PERFORM PROCEDURE B WAS 40 MINUTES

SAMPLE NUMBER 3

O C O C O O O C O O
C 1 O O O O O C O O
C C O O O O 1 1 O O
O C O O O O O O O O
C C O O O O O C O O

AGGLUTINATION IN GP-TWO

AGGLUTINATION IN GP-THREE

TIME TO PERFORM PROCEDURE A WAS 31 MINUTES

TIME TO PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 4

0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 1 0 0 1 0
0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0

AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-THREE
AGGLUTINATION IN GP-FOUR
TIME TO PERFORM PROCEDURE A WAS 33 MINUTES
TIME TO PERFORM PROCEDURE B WAS 45 MINUTES

SAMPLE NUMBER 5

0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0

FALSE POSITIVE IN GP-ONE
AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-FIVE
TIME TO PERFORM PROCEDURE A WAS 30 MINUTES
TIME TO PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 6

0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0

AGGLUTINATION IN GP-THREE
TIME TO PERFORM PROCEDURE A WAS 20 MINUTES
TIME TO PERFORM PROCEDURE B WAS 35 MINUTES

SAMPLE NUMBER 7

C C O O O O O O O O
O C C O O C 1 O O O
O C O O O O O O O O
O C O O O O O C O O
O O O O O O O C O O

AGGLUTINATION IN GP-TWO
TIME TC PERFORM PROCEDURE A WAS 23 MINUTES
TIME TC PERFORM PROCEDURE B WAS 45 MINUTES

SAMPLE NUMBER 8

1 C C O O O O O O O O
O C O C O O O C O O
1 C O O O O O O O O O
C C C 1 1 O O C O O
O O O O O O O O O O

AGGLUTINATION IN GP-ONE
FALSE POSITIVE IN GP-TWO
AGGLUTINATION IN GP-THREE
AGGLUTINATION IN GP-FOUR
TIME TC PERFORM PROCEDURE A WAS 39 MINUTES
TIME TC PERFORM PROCEDURE B WAS 45 MINUTES

SAMPLE NUMBER 9

C C C C C O O 1 O O
O C 1 O O O O O O O
O C C O O C O O O O
O C O O O O O O O O
O O O O O O O O O O

AGGLUTINATION IN GP-ONE
AGGLUTINATION IN GP-TWO
TIME TC PERFORM PROCEDURE A WAS 25 MINUTES
TIME TC PERFORM PROCEDURE B WAS 35 MINUTES

SAMPLE NUMBER 10
 C C O C O O O O O O
 O C O O O O O O O O
 C C O C O O O C O O
 O C O O O O O O O O
 C 1 C O O O O C O O

AGGLUTINATION IN GP-FIVE
 TIME TC PERFORM PROCEDURE A WAS 22 MINUTES
 TIME TC PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 11
 O C C O O C O O O O
 1 C O O O O O O O O
 1 C O C O C C O O O
 O C O O O O O O O O
 C C O 1 O O O O O O

AGGLUTINATION IN GP-TWC
 AGGLUTINATION IN GP-THREE
 AGGLUTINATION IN GP-FIVE
 TIME TC PERFORM PROCEDURE A WAS 36 MINUTES
 TIME TC PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 12
 O C C O O O O C O O
 C C O C O O O O O O
 C 1 C C O 1 O O O O
 O 1 O O O O O O O O
 C C O O O O O O O O

AGGLUTINATION IN GP-THREE
 AGGLUTINATION IN GP-FCUR
 TIME TC PERFORM PROCEDURE A WAS 33 MINUTES
 TIME TC PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 13

0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 1 0
1 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0

AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-THREE
AGGLUTINATION IN GP-FOUR
AGGLUTINATION IN GP-FIVE
TIME TC PERFORM PROCEDURE A WAS 35 MINUTES
TIME TC PERFORM PROCEDURE B WAS 40 MINUTES

SAMPLE NUMBER 14

0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0

AGGLUTINATION IN GP-ONE
TIME TC PERFORM PROCEDURE A WAS 22 MINUTES
TIME TC PERFORM PROCEDURE B WAS 35 MINUTES

SAMPLE NUMBER 15

0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0

AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-FOUR
AGGLUTINATION IN GP-FIVE
TIME TC PERFORM PROCEDURE A WAS 30 MINUTES
TIME TC PERFORM PROCEDURE B WAS 40 MINUTES

SAMPLE NUMBER 16

C C C C C C C C C C
O C C C C C C C C C
C C C C C C C C C C
O C C C C C C C C C
O C C C C C C C C C

AGGLUTINATION IN GP-FIVE
TIME TC PERFORM PROCEDURE A WAS 22 MINUTES
TIME TC PERFORM PROCEDURE B WAS 25 MINUTES

SAMPLE NUMBER 17

C C C C C C C C C C
O C C C C C C C C C
C C C C C C C C C C
1 C C 1 1 C C C C C C
O C C C C C C C C C

FALSE POSITIVE IN GP-ONE
AGGLUTINATION IN GP-FOUR
TIME TC PERFORM PROCEDURE A WAS 29 MINUTES
TIME TC PERFORM PROCEDURE B WAS 35 MINUTES

SAMPLE NUMBER 18

C C C C 1 C C C C C
1 C C C C C C C C C C
O C C C C C C C C C
C C C C C C C C C C
O C C 1 C C C C C C

AGGLUTINATION IN GP-ONE
AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-FIVE
TIME TC PERFORM PROCEDURE A WAS 33 MINUTES
TIME TC PERFORM PROCEDURE B WAS 45 MINUTES

SAMPLE NUMBER 19
C C C C O O C C O O
O C 1 O C O C C O O
C 1 C C O O C C O O
O C O O O O O O O O
O C O C O O O O O O

AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-THREE
FALSE POSITIVE IN GP-FOUR
TIME TO PERFORM PROCEDURE A WAS 39 MINUTES
TIME TO PERFORM PROCEDURE B WAS 35 MINUTES

SAMPLE NUMBER 20
O 1 O C O O O C 1 O
C 1 O O O O C C O O
O C O O O O C O O O
C C O C O O C C O O
O O O O O O C O 1

AGGLUTINATION IN GP-ONE
AGGLUTINATION IN GP-TWO
AGGLUTINATION IN GP-FIVE
TIME TO PERFORM PROCEDURE A WAS 36 MINUTES
TIME TO PERFORM PROCEDURE B WAS 25 MINUTES

APPENDIX 3

This is the basic model in which none of the variables is fixed. Therefore, results with this model should indicate most accurately if there is a significant difference between analysis times for the two procedures.

The calculated Z value of 0.06 requires that the null hypothesis of no difference between mean analysis times for the two procedures be accepted at the .05 level. Thus, it can be concluded that for the ranges of sample contaminants, technician times, level of false positives and number of positives within samples chosen for this demonstration run, we can have 95% confidence in stating that there is no difference between the analysis times required for the two procedures. Or, stated another way, we must conclude that the observed difference between means is due to chance at this level of confidence.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.97249
95% LOWER CONFIDENCE LIMIT 34.09283
95% UPPER CONFIDENCE LIMIT 35.85213
VARIANCE OF ANALYSIS TIME 80.56763

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.93750
95% LOWER CONFIDENCE LIMIT 34.23895
95% UPPER CONFIDENCE LIMIT 35.63603
VARIANCE OF ANALYSIS TIME 50.80859

THE Z STATISTIC FOR MEANS IS 0.06105

APPENDIX 4

In this variation of the basic model the number of contaminants in each sample to be analyzed is held constant. The purpose of this variation is to observe the effect of fixing sample contamination on the calculated Z value. In terms of laboratory application, this models the procedure of performing a large number of analyses on identical samples (samples containing the same number of contaminants). This result clearly can't be obtained with any degree of accuracy in the laboratory and is included to demonstrate the power of simulation techniques such as the model presented.

The calculated Z value of 0.27738 requires that the null hypothesis be accepted but clearly gives a larger Z value than the basic model which indicates that there is a more significant difference between mean analysis times with this variation than with the basic model.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 35.48749
95% LOWER CONFIDENCE LIMIT 34.81630
95% UPPER CONFIDENCE LIMIT 36.15866
VARIANCE OF ANALYSIS TIME 46.90576

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 35.34999
95% LOWER CONFIDENCE LIMIT 34.64754
95% UPPER CONFIDENCE LIMIT 36.05243
VARIANCE OF ANALYSIS TIME 51.37817

THE Z STATISTIC FOR MEANS IS 0.27738

APPENDIX 5

In this variation, the probability of a "false positive" was set at zero. The result is as might be anticipated, in that the number of samples analyzed under procedure A is reduced and the analysis time is shortened considerably.

The negative Z value indicates that the times for procedure B were greater than the times for procedure A and, the hypothesis of no difference between mean analysis times is rejected with the calculated Z of -5.71084. Thus, under the conditions of this demonstration it can be concluded with 95% confidence that there is a difference between means and, because the Z value is negative, that procedure A is significantly better than procedure B. In fact, referring to the Normal probability tables, it can be seen that with a Z value this large our confidence can exceed 99%. A sensitivity analysis was performed with the following results:

<u>Probability of a "False Positive"</u>	<u>Z Value</u>
.1	-2.28
.11	-2.02
.111	-1.99
.112	-1.93

Thus, the critical value of probability for false positives is slightly less than .112, that is, as the probability of a false positive approaches .111 from above, the Z value reaches the point (-1.96) at which the hypothesis must be rejected.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 31.84999
95% LOWER CONFIDENCE LIMIT 31.05318
95% UPPER CONFIDENCE LIMIT 32.64679
VARIANCE OF ANALYSIS TIME 66.10791

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.93750
95% LOWER CONFIDENCE LIMIT 34.23895
95% UPPER CONFIDENCE LIMIT 35.63603
VARIANCE OF ANALYSIS TIME 50.80859

THE Z STATISTIC FOR MEANS IS -5.71084

APPENDIX 6

In this variation, the analysis time for a technician to examine one group under procedure A was fixed at seven minutes per positive group. As expected, the variance dropped from 80 plus with the basic model to 59.66992 with this model. This is an indicator of the overall contribution of variation in technician time (between technicians) to the variance of the procedure. No significant difference in the Z value is observed.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.79250
95% LOWER CONFIDENCE LIMIT 34.03548
95% UPPER CONFIDENCE LIMIT 35.54950
VARIANCE OF ANALYSIS TIME 59.66992

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.93750
95% LOWER CONFIDENCE LIMIT 34.23895
95% UPPER CONFIDENCE LIMIT 35.63603
VARIANCE OF ANALYSIS TIME 50.80859

THE Z STATISTIC FOR MEANS IS -0.27591

APPENDIX 7

In this variation, the analysis time for a technician on procedure B was fixed at seven minutes per group. As expected, the variance in results for procedure B dropped to zero. This serves as a further check of the validity of the program.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.97249
95% LOWER CONFIDENCE LIMIT 34.09283
95% UPPER CONFIDENCE LIMIT 35.85213
VARIANCE OF ANALYSIS TIME 80.56763

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 35.00000
95% LOWER CONFIDENCE LIMIT 35.00000
95% UPPER CONFIDENCE LIMIT 35.00000
VARIANCE OF ANALYSIS TIME 0.00000

THE Z STATISTIC FOR MEANS IS -0.06130

APPENDIX 8

As a final check on the operation of the computer program with parameters fixed, both technician times were fixed. The results confirm those obtained in appendices 6 and 7 for variances of the two procedures. Further, the Z value of -0.53725 remains in the acceptance range, further demonstrating the effect of technician time between the two procedures. These could be considerably more significant in a situation where there were either more technicians involved in the procedures or where the variability between individual technician times was greater.

COMPUTER OUTPUT

RESULTS FOR PROCEDURE A

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 34.79250
95% LOWER CONFIDENCE LIMIT 34.03548
95% UPPER CONFIDENCE LIMIT 35.54950
VARIANCE OF ANALYSIS TIME 59.66992

RESULTS FOR PROCEDURE B

NUMBER OF SAMPLES ANALYZED 400
MEAN OF ANALYSIS TIME WAS 35.00000
95% LOWER CONFIDENCE LIMIT 35.00000
95% UPPER CONFIDENCE LIMIT 35.00000
VARIANCE OF ANALYSIS TIME 0.00000

THE Z STATISTIC FOR MEANS IS -0.53725

APPENDIX 9

The objective of this appendix is to present a general discussion of cost analysis as it might be applied to the question of choosing between laboratory procedures based on total cost. Specifically, applications of data obtained with the simulation model to cost analysis will be discussed. Further, because computer facilities may not be readily available to the laboratory, mathematical estimation procedures which may be employed without the simulation model will be presented.

Costs associated with the laboratory procedures of interest will be categorized and discussed individually. A model for treating the uncertainty associated with these costs will be described. Categorization is an important step in preparing a cost analysis and should not be skipped over lightly. One sure way to minimize cost in any analysis is to overlook or purposely omit some relevant cost. The decisionmaker should not permit this to happen without good justification. A laboratory supervisor can easily obtain a precise and reliable estimate of some of the costs of a laboratory procedure. That data alone, however, is not really helpful in many instances. It is very difficult to make a rational choice between proposed laboratory procedures A and B, no matter how detailed and precise and dependable the cost figures, if the figures represent only some uncertain fraction of the total analysis cost of each

procedure. The decisionmaker needs to compare, as well as he can, their respective total costs.

Thus, the real challenge facing the individual preparing a cost analysis is to be as comprehensive as possible in the analysis. Because there are a few readily identifiable costs that can be conveniently identified, measured, and evaluated, we focus attention on these and give little, if any, attention to those costs that are less easily identified measured and evaluated.

Clearly, there is a difference between dollar expenditures during a period of time and total cost during that same period. If the laboratory supervisor is limiting his analysis to that portion of cost associated directly with immediate dollar outlay, this cost might well be labeled "dollar expenditure" rather than "total cost". Most costs can, at some point, be translated either into dollar expenditures or expenditures of resources that can be evaluated in terms of dollars. However, there is another category of costs that fall into neither of the above dollar categories. This includes such intangibles as "convenience", "acceptability" and the like. Clearly, these must be taken into consideration by the laboratory supervisor but for purposes of this discussion on cost analysis, these intangibles will be ignored.

Generally, the laboratory supervisor is required to perform cost analyses on procedures in operation for budgetary or other administrative purposes. However, cost

analysis is also indicated when the cost of equipment and reagents is sufficiently high to warrant an investigation of the trade-off between total analysis time and total analysis cost.

Clearly, the procedure that requires significantly less analysis time, costs less to perform and provides an acceptable level of reliability is the procedure to select! On the other hand, when expendable costs associated with a procedure is low, it seems reasonable to select those procedures which require less analysis time as in the example presented with the simulation model.

Our primary interest is in examining those procedures which pose a question regarding the additional cost associated with saving analysis time. Or, stated another way, how much additional analysis time will we expend in order to save dollar costs. Finally, since our other variable, time, also costs money in the laboratory we must aggregate time with other cost considerations previously mentioned into one workable model and solve the problem:

Minimize: Cost of Analysis

Subject to: Reliability Constraints

In most laboratory procedures, the question of reliability is dealt with first. More precisely, most laboratory supervisors will not be faced with the problem of selecting between procedures which do not meet a minimum level of reliability. This is especially true if the laboratory is engaged in contractual quality control work for which most

of the laboratory procedures are rather clearly spelled out in contractual publications. Therefore, the laboratory supervisor need only examine the question of minimizing cost.

Laboratories wishing to use cost analysis as a decision tool will generally fall into one of the following categories:

1. Case 1 - The laboratory has been performing a procedure routinely for an extended period of time and has decided to consider an alternative (but similar) procedure. In this case, the cost analysis will be fairly straightforward because the laboratory can use data on hand from the current procedure and either simulate or estimate by direct mathematical means the relevant parameters for the new procedure.

2. Case 2 - The laboratory is interested in selecting the most cost efficient of two procedures which have not been performed in the laboratory on a routine basis. In this case, data relevant to these procedures will not be readily available to the analyst and must, therefore, either be obtained from an outside source (such as another laboratory) or collected experimentally in the laboratory.

The value of data obtained from another laboratory may be of questionable value unless the analyst has first hand knowledge of the circumstances surrounding the collection and compilation of the data. Because there is normally a great number of areas in which laboratories differ, the use

of data obtained from outside laboratories must rank very low in the order of preference for data sources.

A preferable approach, if resources permit, is to perform both procedures on an experimental basis in the laboratory, collect data and base decisions on that data. If it is impractical to perform both procedures on an experimental basis, as is often the case, then simply select one of the procedures on an intuitive basis and use it for a reasonable period. When sufficient data is available, either model the second procedure using data obtained from the first and/or estimate parameters mathematically based on data from the first. In any case, it seems reasonable that data collected in the laboratory by making direct observations of the personnel and laboratory environment in question is preferable to using data obtained in another laboratory with different personnel working in a different environment.

The point is that results obtained with either a simulation model or a direct analytic model are no better than the data entering the model. Therefore, as much care as seems appropriate should be exercised in choosing the data base for a cost analysis.

Data Base

In order to make this discussion relevant to the type of procedures under consideration in the simulation model, all cost data will be discussed in terms of the positive group

unit. At the same time, the general approach to be employed in this presentation is equally applicable in most respects to laboratory procedures for which the sample unit is not readily divisible into identifiable groups or subgroups.

The first step in preparing a cost analysis for these procedures is to categorize the costs associated with these procedures. Keeping in mind the basic requirement that costs be categorized as comprehensively as seems appropriate to the procedures in question, the following cost categories are established:

Cost Categories

<u>Variable</u>	<u>Direct</u>	<u>Indirect</u>
Time related	1. Technician 2. Facilities	1. Storage loss 2. Samples not tested
Positive group related	1. Reagents 2. Glassware 3. Equipment Maint. and calibration	1. Procurement and supply
<u>Fixed</u>	1. Reporting 2. Clerical	1. General Admin. 2. Overhead (Janitorial, utilities, etc.)

Step two in the costing process is to obtain values for each cost input and then to combine the individual input costs into the appropriate variable and fixed cost categories shown in the table above. If the analyst has constant or very predictable values for each input in a cost category, then the individual input costs need only be added together to obtain a category cost value. The term "very predictable" in this context is used to describe a value for which the variance is insignificant or has been accurately established by some reliable means.

Generally, the individual costs in each category are neither constant nor very predictable and, therefore, it is necessary to consider the question of uncertainty associated with each input in the cost analysis.

Although most of the individual inputs in each of the categories of variable and fixed costs are self explanatory, a brief discussion of the cost estimating aspects of each and an approach to the question of treating uncertainty follows.

To the laboratory supervisor who is not firmly grounded in probability and statistical theory, the question of treating uncertainty in a cost analysis of this type may seem overwhelming. The unfortunate result is that a cost model which ignores uncertainty is often employed. Clearly, what is required is a model which permits the laboratory supervisor to improve cost estimates by considering uncertainty associated with inputs and, at the same time, does

not require an unrealistic investment in data collection or statistical analysis for each input parameter.

One model which fits this basic criteria is presented in a Rand technical publication (6). This model requires that the analyst know only the lowest possible, most likely and highest possible (denoted by L, M and H) values for each input parameter to be used in the model. Further, it must be assumed that there is a ten percent probability of the actual value being lower than L and a ten percent probability of the actual value being higher than H. Then, a simple approximation of the expected value or mean becomes

$$\bar{X} = \frac{X_L + 4X_M + X_H}{6}$$

and, employing the assumptions above, the range $X_H - X_L$ varies between 2.5 and 2.9 standard deviations for a wide class of distributions including rectangular, exponential, triangular, normal and beta. Thus we write

$$X_H - X_L = 3\sigma_X$$

where σ is the standard deviation. Then,

$$\sigma_X = \frac{X_H - X_L}{3}$$

Application of this model to the cost categories listed in step one is as follows:

A. TIME RELATED COSTS

Obtain values of L, M and H for each of the costs in this category and denote each as shown in the individual variable sections below.

1. Technician - Denote L, M and H as b_{1L} , b_{1M} and b_{1H} . These values can be obtained from personnel or finance offices for each technician and then a weighted average calculated for b_{1M} .

2. Facilities Utilization - Denote L, M and H as b_{2L} , b_{2M} and b_{2H} . For most laboratory procedures, the facilities utilization costs include such items as laboratory bench space, associated instrumentation, holding facilities, incubation facilities and the like.

3. Storage Loss - Denote these as b_{3L} , b_{3M} and b_{3H} . Costs in this item are those resulting from holding or storing quantities of the product while laboratory analysis is in progress. That is, the additional storage costs incurred by the delay in obtaining laboratory results.

4. Samples Untested - Denote these as b_{4L} , b_{4M} and b_{4H} . These costs refer to loss and/or deterioration of product held for which testing is not accomplished due to utilization of laboratory resources for other testing procedures.

Now, although we have no real idea of the exact shape or characteristics of the time related cost distribution which we are attempting to describe, the expected value (mean) and standard deviation may be estimated by the following:

$$\text{Let } b_L = \sum_{i=1}^4 b_{iL}$$

$$b_M = \sum_{i=1}^4 b_{iM}$$

$$b_H = \sum_{i=1}^4 b_{iH}$$

Then, the mean is

$$\bar{b} = \frac{b_L + 4b_M + b_H}{6}$$

and the standard deviation is

$$\sigma_b = \frac{b_H - b_L}{3}$$

B. POSITIVE GROUP RELATED COSTS

Obtain values of L, M and H for each of the costs in this category and denote each in a manner similar to that for time related costs.

1. Reagents - Denote these as c_{1L} , c_{1M} and c_{1H} . On a per positive group basis, the variance associated with these costs should be reasonably small and, therefore, should not be a real problem to estimate.

2. Glassware - Denote these as c_{2L} , c_{2M} and c_{2H} . This cost item is intended to include preparation, handling, replacement and loss resulting from the analysis of a positive group. In general, it should also include those items of cost resulting from preparation and handling of all appliances and utensils employed in the procedure.

3. Equipment - Denote these as c_{3L} , c_{3M} and c_{3H} . This item is intended primarily to include those maintenance and calibration costs associated with balances, recorders and similar equipment which result directly from the performance of the laboratory procedure in question.

4. Procurement and Supply - Denote these as c_{4L} , c_{4M} and c_{4H} . This item is self explanatory but might be one of the more difficult to estimate.

$$\text{Now, let } c_1 = \sum_{i=1}^4 c_{iL}$$

$$c_2 = \sum_{i=1}^4 c_{iM}$$

$$c_3 = \sum_{i=1}^4 c_{iH}$$

Then the mean is

$$\bar{c} = \frac{c_L + 4c_M + c_H}{6}$$

and the standard deviation is

$$\sigma_c = \frac{c_H - c_L}{3}$$

C. FIXED COSTS

Unlike the two categories above, fixed costs will be on a per sample basis. Further, because the relative variance associated with these costs is small compared to the variances associated with the two categories above, these costs might be treated as constants.

1. Reporting - Denote this as a_1 .

The process of reporting on most analytic procedures of interest in the laboratory consists of entering raw data on a standard reporting form and delivering it to the administrative office for further processing. Therefore, the between sample variance should not be too great.

2. Clerical - Denote this as a_2 .

Typing reported results from analyses in the laboratory is a fairly standard procedure and, clearly, it requires no

more effort to type 1,000 MPN than to type 100 MPN. Perhaps I should say very little more effort! At any rate, the variance should be small for this item and it probably should be treated as a constant.

3. General Administrative - Denote this as a_3 .

This indirect cost is not time related or positive group related and can easily be divided equally between samples analyzed. Again the variance should be small.

4. Other Overhead - Denote this as a_4 .

The procedures under consideration in this model require variable amounts of total analysis time and, since overhead cost is related to time utilized in each procedure, it might be reasonable to allocate a fixed portion of overhead such as utilities, janitorial services and the like to each sample analyzed on the basis of a total fraction of laboratory time required to perform each procedure. For example, if the laboratory has five full time technicians and operates on a 40 hour week basis, the laboratory then has 200 analysis hours available. If the procedure in question requires a total of 20 analysis hours weekly, then allocate one tenth of other overhead costs to this procedure. Divide the amount allocated to this procedure by the number of samples analyzed and treat this as the cost per sample of other overhead.

$$\text{Now, let } a = \sum_{i=1}^4 a_i$$

and treat a as a constant in the analysis.

Now, having obtained an estimate for each applicable cost category and, acknowledging that there is considerable uncertainty associated with most of these estimates, they may be aggregated as follows:

$$\text{Total Cost/Group} = \text{Fixed Cost/Gp.} + \text{Variable Cost/Gp.}$$

$$\text{Then, Expected Total Cost} = a + \bar{b}\bar{x} + \bar{c}\bar{y} = f$$

where a = Fixed Cost

\bar{b}, \bar{c} = Mean Cost/unit (i.e. dollars/hour or per Pos. Gp.)

\bar{x}, \bar{y} = Variable No. Units (time or Pos. Gps.)

$$\begin{aligned} \text{Then, Variance of Cost} = & \left[f_x(\bar{x}, \bar{y}, \bar{b}, \bar{c}) \sigma_x \right]^2 + \left[f_y(\bar{x}, \bar{y}, \bar{b}, \bar{c}) \sigma_y \right]^2 \\ & + \left[f_b(\bar{x}, \bar{y}, \bar{b}, \bar{c}) \sigma_b \right]^2 + \left[f_c(\bar{x}, \bar{y}, \bar{b}, \bar{c}) \sigma_c \right]^2 \end{aligned}$$

as an approximation where f_x means derivative of f with respect to the variable x .

With this estimate of the mean and variance of total cost for each of the two procedures in question, it is possible to perform hypothesis testing and determine if there is a significant difference between the expected costs for the two procedures. In the calculations above, it should be noted that in those instances where the variance of one variable is small compared to the variance of a variable by which it is being multiplied, then the variable with the smaller variance can be treated as a constant and the computations thereby greatly simplified.

As shown in appendices 3-8, both the means and variances for the variables x and y are readily obtained from the simulation model. In the laboratory not having access to a simulation model such as this, these values may be estimated

(roughly) from either existing laboratory data or from experimental work done in the laboratory. In either case, the following mathematical approach may be used in estimating x and y using only the expected value of input parameters.

DIRECT ESTIMATE USING MEANS

1. Actual

The probability of a microorganism entering a group on the first trial is 1/5. Then on each succeeding trial, probability statements must be based on the conditional probabilities resulting from the first trial. This procedure gets very complicated after only a few trials.

2. Estimate

Using the same initial probability of a microorganism entering a group (1/5) and, applying the binomial distribution for an average (mean) number of contaminants per sample of three, the probability that a sample contains one or more contaminants in one or more groups becomes

$$\sum_{i=1}^3 \text{Probability (Number Positive Groups = i)}$$

Let $p = \text{Probability of Positive Group} = 1/5$

$$q = 1 - p = 4/5$$

Then, in three trials (3 contaminants/sample)

$$P(0 \text{ Contaminants in a Group}) = \frac{3!}{0!(3-0)!} (.2)^0 (.8)^3 = .512$$

$$\text{Thus, } P(\text{Contaminant in a Group}) = 1 - .512 = .488$$

or, about .5 of Groups are positive (≈ 2.5 Gps.). Add this to the probability of a false positive (.2) or, on the

average of 1 of 5 groups \approx 1 group/sample, then average number of Positive Groups/Sample = $3.5 = \bar{y}$.

For Procedure A:

Setup time = 15 minutes (Average)

Positive Groups = 3.5 (Including False Positives) = \bar{y}

Average Tech. Time = 7 min/group = b

Total Analysis Time = $39.5 = \bar{x}_A$

and, for Procedure B:

Total Analysis Time = $5 \times 7 = 35 = \bar{x}_B$

From Model (for comparison)

Procedure A = $34.97 = \bar{x}_A$

Procedure B = $34.93 = \bar{x}_B$

Finally, it should be recalled that total analysis costs may change with time and quantity of samples analyzed. Most laboratory personnel are familiar with the improved efficiency that normally results from experience with most laboratory procedures. In general, this improved efficiency can be thought of as a "learning curve" effect.

Further, because the rate at which learning occurs with one procedure may be significantly different than the rate at which learning occurs with another procedure, it follows that costs evaluated on the basis of a few experimental sample lots may be significantly different than costs evaluated on comparable sample lots when the learning effect is taken into consideration.

Because the learning curve effect is a significant factor which should be included in a cost analysis approach

to selecting the most efficient laboratory procedure, the final sections of this appendix will contain a discussion of the theory and practice of learning curves. This discussion is intended to be comprehensive enough for application to the problem at hand. For a more complete treatment of the subject, the reader is referred to the Rand Publication (6) from which most of this material is taken.

THEORY OF LEARNING CURVES

The basis of learning curve theory is that each time the total quantity of items produced (samples analyzed) doubles, the cost per item (sample) is reduced to a constant percentage of its previous cost. Alternative forms of the theory refer to the incremental (unit) cost of producing an item at a given quantity or to the average cost of producing all items up to a given quantity. For example, if the cost of analyzing the 200th sample is 80 percent of the cost of analyzing the 100th sample, and if the cost of the 400th sample is 80 percent of the cost of the 200th and so forth, the process of analyzing samples is said to follow an 80 percent unit learning curve. If the average cost of analyzing all 200 samples is 80 percent of the average cost of analyzing the first 100 samples, the process follows an 80 percent cumulative average learning curve.

Either formulation of the theory results in a power function that is linear on logarithmic grids. Figure 1 shows a unit curve for which the reduction in cost is 20 percent with each doubling of cumulative sample output.

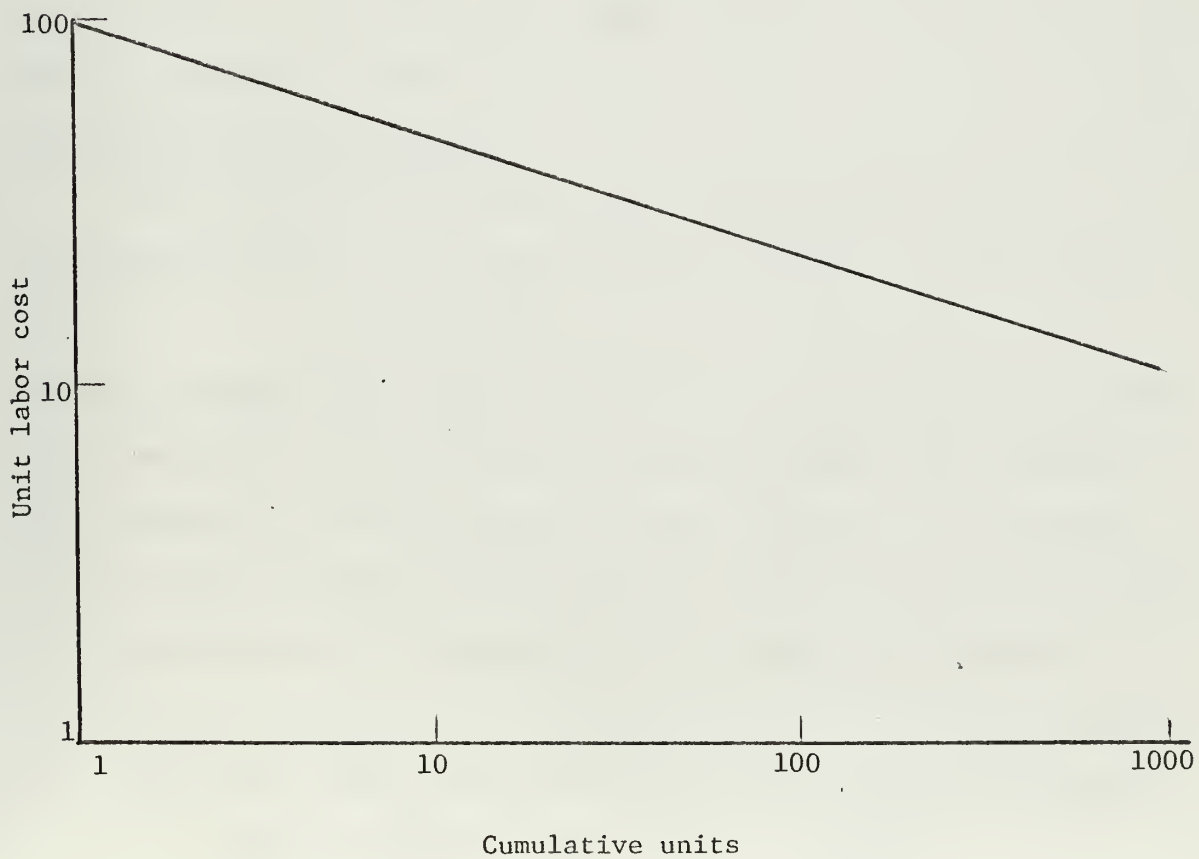
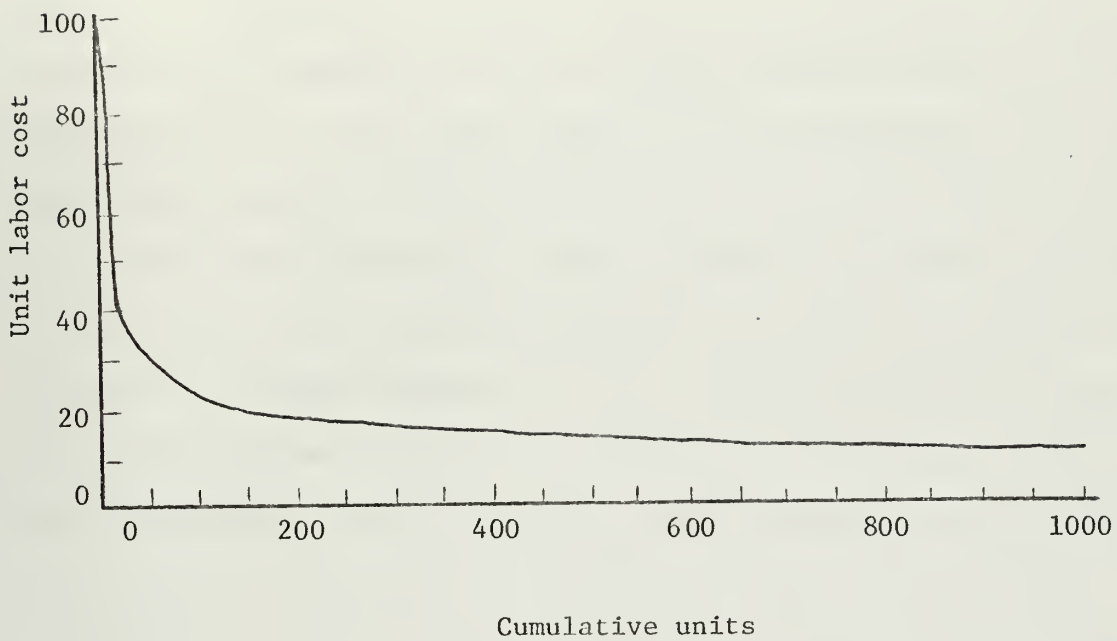


Figure 1 - The 80 percent learning curve on arithmetic and logarithmic grids

The upper figure shows the curve on arithmetic grids and the lower on logarithmic grids. The arithmetic plot shows the percentage reduction in cost in each sample analyzed is very pronounced for the early units. On an 80 percent curve, for example, cost decreases to 28 percent of the original value over the first 50 units. Over the next 50 samples analyzed, it declines only 5 more percentage points, i.e., down to 23 percent of sample number 1 cost. The factors that account for the decline in unit cost as cumulative output increases are numerous. Obviously, one major contribution is due to task familiarization by technicians which results from repetition of the analytic procedures. Many of the other factors are not clearly understood and no attempt will be made to enumerate them here.

The Log-Linear Hypothesis

The relationship between cost and quantity may be represented by a power (log-linear) equation of the form

$$y = ax^b$$

where x equals the cumulative quantity of samples analyzed. The constant a is the cost of analyzing the first sample. The exponent b, which measures the slope of the learning curve bears a simple relationship to the constant percentage to which the cost is reduced as the number of samples analyzed is doubled. If S represents the fraction to which cost decreases when quantity doubles, the equation becomes

$$S = \frac{y_{2x}}{y_x} = \frac{a(2x)^b}{ax^b} = 2^b \quad \text{or} \quad b = \frac{\text{Log } S}{\text{Log } 2}$$

This equation shows that for a value of S equal to 75 per-
cent, the corresponding value of b is

$$\frac{\text{Log } .75}{\text{Log } 2} \quad \text{or} \quad -.415$$

Plotting a Curve

In the graphical display of learning curves, the problem is to represent the average cost for a lot since, typically, analysis times or costs are not recorded by sample unit.

See, for example, the following table:

<u>Lot</u>	<u>Sample Units</u>	<u>Analysis time per lot in minutes</u>
1	1-10	583
2	11-20	437
3	21-50	1,055
4	51-100	1,475

To plot a cumulative average curve from these data, the cumulative average hours are computed at the final unit in each lot:

<u>Plot Point</u>	<u>Analysis time per lot (min.)</u>	<u>Computation</u>	<u>Cumulative Average Minutes</u>
10	583	583/10	58.3
20	437	1,020/20	51.0
50	1,055	2,075/50	41.5
100	1,475	3,550/100	35.5

The cumulative average at the 10th sample unit is 58.3 minutes; this is the first plot point. Successive plot points are at the end of each lot since these are the points where the cumulative average minute figures apply.

To plot the unit curve it is first necessary to compute the unit minutes and then to establish plot points. The unit minutes can be taken as an average for each lot:

<u>Lot</u>	<u>Computation</u>	<u>Unit Minutes</u>
1	583/10	58.3
2	437/10	43.7
3	1,055/30	35.2
4	1,475/50	29.5

The lots can be represented by these unit hour values. The question is, where should the values be plotted? To plot at the lot arithmetic midpoint is to assume that the learning curve can be approximated by a linear curve on arithmetic grids, but as suggested by Figure 1 such a method of approximation only becomes reasonable for lots following a large number of previous samples. Thus, when dealing with a log-linear function, the arithmetic midpoint plot produces the unequal distribution of the area under the curve as shown in Figure 2.

The true midpoint is defined as that unit, x_m , which represents the entire lot and which must also reflect the average unit cost, y_m , of the lot. The total cost of the lot is equal to the product of y_m and the number of samples in the lot, n . This product will approximate the area under the curve for n units (see Figure 3).

In practice, the mathematics associated with determining actual plot points makes the procedure difficult. Therefore

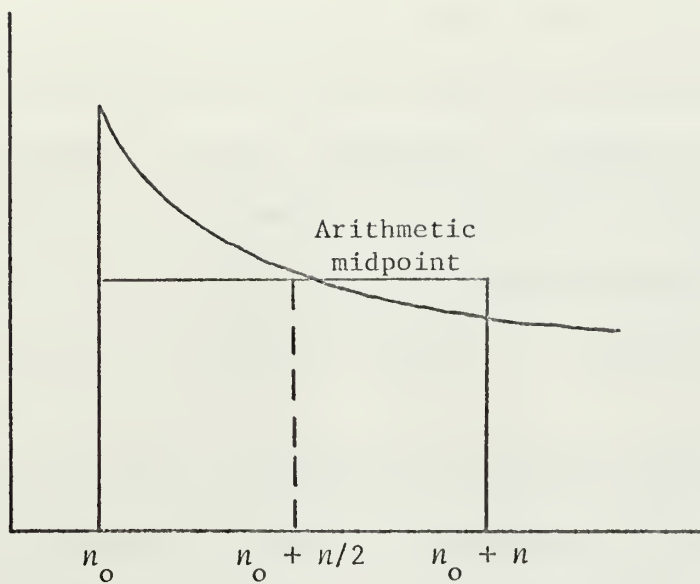


Figure 2 - Learning curve on arithmetic grids

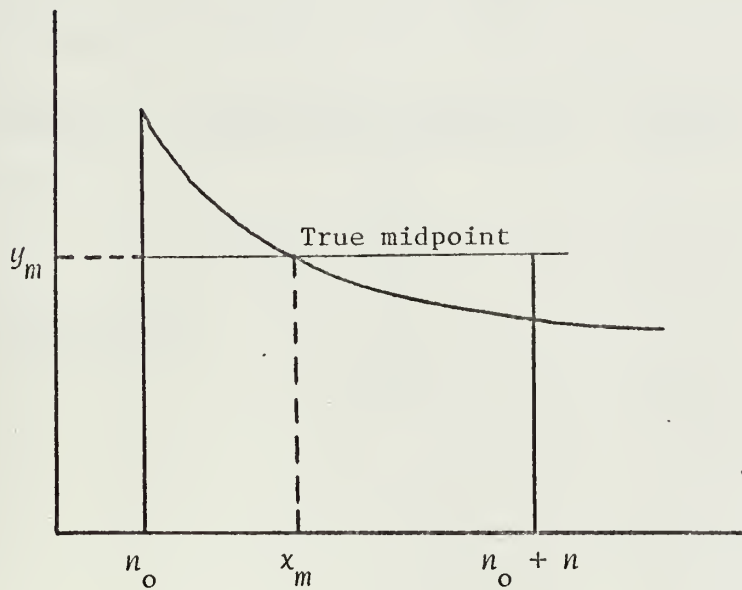


Figure 3 - True lot midpoint on arithmetic grids

when dealing with first few lot quantities which comprise more than about 25 samples, plot points can be taken from graphs provided in the Rand Publication referenced earlier. Or, if graphs are not available, estimate the plot points by computing the arithmetic lot midpoint and then moving it slightly to the left. For succeeding lots, the arithmetic lot midpoint is usually adequate. Consider the following example: -

If the unit and cumulative average curves are plotted as shown on Figure 4, then, to determine the learning rate, simply select two cumulative quantities such that the second is two times as large as the first, read their respective costs from the graph and determine the ratio of the respective costs.

	<u>Curve</u>	<u>Cumulative Quantity</u>	<u>Cost</u>	<u>Learning Rate</u>
1.	Unit	10	5	4.1/5 or 82%
		20	4.1	
2.	Cumulative Average	10	6	5.1/6 or 85%
		20	5.1	

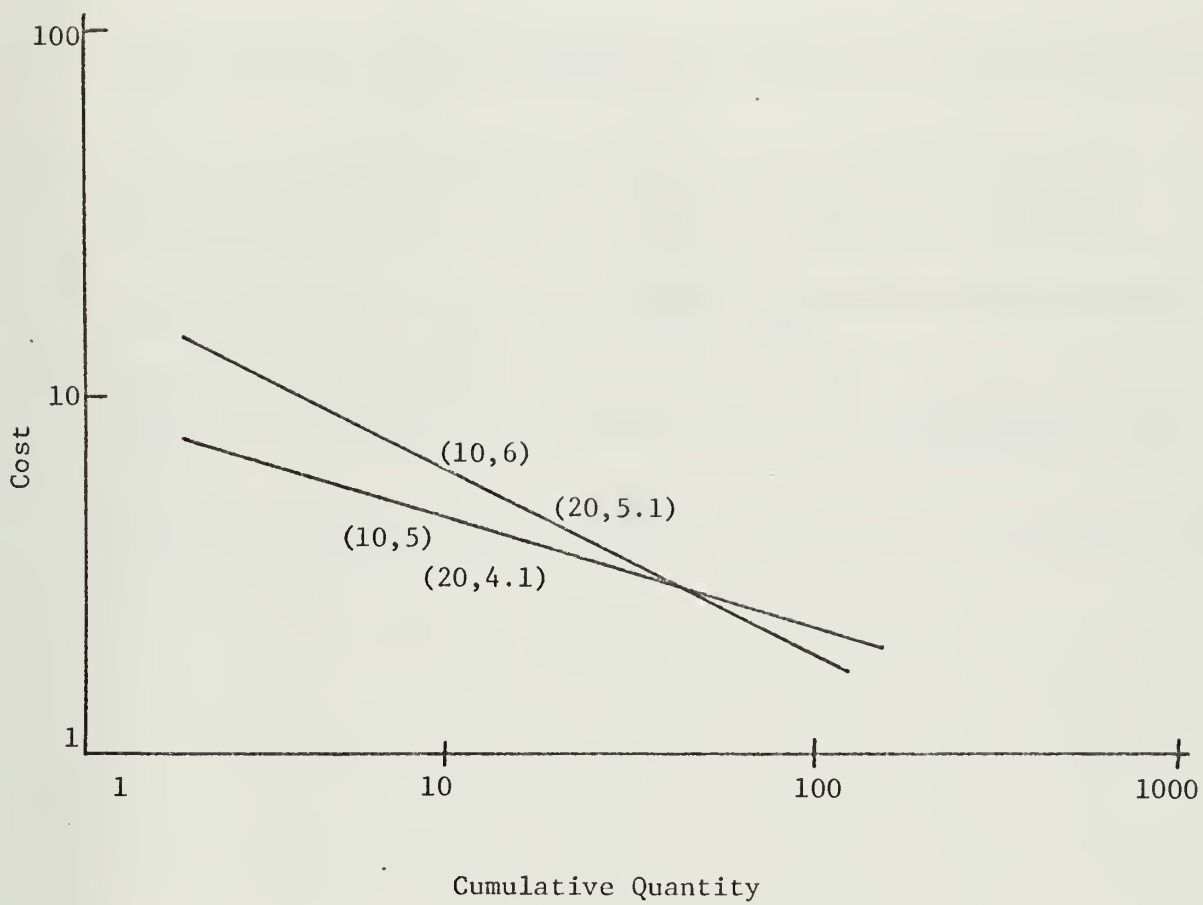


Figure 4

BIBLIOGRAPHY

1. Hall, Herbert E., "Examination of Foods for the Coliform Organism," Training Course Manual, Robert A. Taft Sanitary Engineering Center, Cincinnati, Ohio, April, 1966.
2. Naylor, T.H., Balintfy, J.L., Burdick, D.S., and Chu, K., Computer Simulation Techniques, New York: John Wiley & Sons, Inc. 1966.
3. Halvorson, H.O., and Ziegler, N.R., "Application of Statistics to Problems in Bacteriology," Journal of Bacteriology, Vol. 25, p. 101-121, 1933.
4. U.S. Department of Health Education and Welfare Public Health Service Publication 33, 1946.
5. Cochran, W.G., "Estimation of Bacterial Densities by Means of the "Most Probable Number"," Biometrics, June 1950.
6. The Rand Corporation, "Military Equipment Cost Analysis," June, 1971.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2. Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3. Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
4. Assoc, Professor H. J. Zweig, Code 80 Zw Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
5. Assoc. Professor M. G. Sovereign, Code 64 Zo Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
6. Headquarters, Department of the Army (DASG-VSS) Washington, D.C. 20314	1
7. Major T. S. Armstrong, USA 3805 Atlas Avenue El Paso, Texas 79904	5
8. Chief of Naval Personnel Pers 11b Department of the Navy Washington, D.C. 20370	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)		2a. REPORT SECURITY CLASSIFICATION	
Naval Postgraduate School Monterey, California 93940		Unclassified	
		2b. GROUP	
3. REPORT TITLE			
A Simulation Model of Microbiological Laboratory Procedures			
4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)			
Master's Thesis; September 1973			
5. AUTHOR(S) (First name, middle initial, last name)			
Tommy S. Armstrong			
6. REPORT DATE		7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1973		86	6
8a. CONTRACT OR GRANT NO.		9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO.			
c.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. DISTRIBUTION STATEMENT			
Approved for public release; distribution unlimited			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY	
		Naval Postgraduate School Monterey, California 93940	
13. ABSTRACT			
<p>Laboratory procedures, mathematical theory and distribution assumptions associated with two microbiological testing techniques are presented. A computer simulation model is then formulated and programmed based on these procedures, and thus the influences of changes in the number of microorganisms per sample, distribution of microorganisms within the sample, number of positive groups, probability of "false positives", distribution of "false positives" and technician analysis times are determined.</p> <p>Using the basic simulation model as an experimental device, an example is presented to demonstrate its use in estimating the total time required to analyze a sample using each of the two procedures. Five variations of the basic model are presented to demonstrate the model's flexibility and sensitivity to fixing individual parameters.</p> <p>Hypothesis testing is conducted on data obtained with the basic model and five variations. A significant Z value was obtained with variation two in which the probability of a false positive was set at zero. Results of all hypothesis testing are presented and a discussion of model data application in cost analysis is appended.</p>			

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT



Thesis
A7156
c.1

Armstrong

145934

A simulation model of
microbiological labora-
tory procedures.

Thesis
A7156
c.1

Armstrong

145934

A simulation model of
microbiological labora-
tory procedures.

thesA7156

A simulation model of microbiological la



3 2768 001 00637 2

DUDLEY KNOX LIBRARY